Sarah F. Majors

(724) 683 - 5500
https://www.linkedin.com/in/sarah-f-majors/

sfmajors373@gmail.com
www.github.com/sfmajors373

OVERVIEW

Originally an archaeologist, I have transitioned to tech and have spent my first tech jobs working in web development with a side of data analysis and data visualization. Now I am looking to obtain a job in artificial intelligence.

MACHINE LEARNING PROJECTS

Breast Cancer Capstone

- Neural network was designed and trained against stained lymph node slides, from the Camelyon 16 dataset, to perform image segmentation with the goal of the detection of cancerous and abnormal cell structures
- Various methods such as OTSU were used to efficiently clean, tile, and mask the slide data to reduce man hours necessary for production of a dataset as well as enhance the quality of the training set and model accuracy
- Convolutional Neural Network model was designed to flag images from Camelyon 16 dataset as abnormal
- UNet based image segmentation was performed agaisnt Camelyon 16 dataset images identified by the CNN
- Pix2Pix Generative Adversarial Network was used to expand the existing Camelyon 16 dataset to create more training data to feed the pipeline to enhance reliablity and accuracy by balancing positive and negative samples
- Both GPU and CPU computation were utilized to reduce training time and increase the velocity of development
- Set up and deployed a Jupyter Labs based research environment in a Docker container based on Arch Linux on a 32 core Threadripper machine with a mirror of the Camelyon 16 dataset for each team member to reduce the set up time and allow for rapid iterations
- Deployed the model in a flask application on Google Cloud allowing a user to upload a whole slide or just an image tile and see the prediction as well as the segmentation, if applicable

Privacy Preserving Satellite Imagery   *in progress*

- Produced a demo for Open Mined to implement a machine federated learning model in which data owners have control of the data and the data scientists never access the raw data thereby keeping the data private whilst still allowing the end user to have meaningful results
- The model is trained on satellite imagery to locate pools in certain areas as a proof of concept that these methods of data privacy can be used on data of this nature
- Created reusable Docker container hosting Jupyter Labs and all necessary machine learning libraries to reduce issues related to dependency management
- Compiled PySyft and PyGrid packages for Arch Linux and became the maintainer of those and several other machine learning packages in the Arch User Repository
- Built computer to reduce dependency on expensive cloud compute and prevent runaway costs while experimenting with various models, parameters, etc as well as speed up training time

Chest X-Ray Pathology Detection

- Utilized the CheXpert dataset to develop a potential low cost solution to detect chest pathologies, such as cardiomegaly, edema and pleural effusion, deployed to Raspberry Pi via Flask with three other people
- Trained various models, such as VGG and Resnet, to achieve best results in given time, with VGG16 having the best results at 79.15% accuracy rate with a 50,000 image training set
- Created a small flask application to deploy on Raspberry Pi to upload images and make pathology predictions
- Recorded video demo of the project for hackathon submission

Federated Learning Cluster

- Created a federated learning computational cluster on a set of four Raspberry Pi 3s and 4s to have a platform to further explore how federated learning works
- Built an automated tool to create Arch Linux ARM rootfs for Raspberry Pi 3s and 4s using pacman and bash
- Practiced compiling Arch packages for ARM so I could use Arch on the Pis

Noise Analysis on Android

- Worked with three others to compare accuracy of local and global differential privacy using federated learning via the PySyft library
- Utilized Gaussian and Laplacian curves as the Lapace mechanism to analyze how the difference in noise impacted accuracy, in addition to testing without noise
- Tested in an Android environment using the Synthetic Digits dataset and created outcome visualizations using GraphView

WORK HISTORY

Teamsense *January 2022 - June 2022*
- Contribute to implementation design of new features or architecture as well as implement new features, maintain existing code, and correct defects
- Take part in rotation to act as support engineer for the customer success team to triage bugs, oversee their resolution and ensure that defects are corrected in accordance with the contract with the client
- Tested and compared new hardware options for devs to be able to work more efficiently with fewer restrictions due to compute power
- Researched various dashboard software options, contacted representatives from the ones most closely suiting company needs and presented the information to the team
- Tech Stack: Python, Django, Typescript, React, Datadog, git

Rivers Agile *April 2019 - January 2022*
- Converted existing code to VueJS in order to add internationalization features to prepare for use in Europe
- Chased bugs, created features, and developed prototypes for real time dashboards and performance reports to showcase the efficiency of the autonomous forklifts to the clients with the goal of aiding in sales
- Participated in planning meetings for new software projects in which there was integration with other teams in order to fully understand customer needs from the product and then created user stories from that perspective
- Participated in architectural planning meetings for new projects and tested different technologies by creating rapid prototypes and discussing the pros and cons of different approaches
- Tech Stack: Python, Plot.ly, Django, RabbitMQ, MQTT, VueJS, node, selenium, i18n, Docker, Vagrant, Gitea

Contractor *February 2018 - Jauary 2019*
- Collected data from the FAA pertaining to yearly flight hours per aircraft model and from the NTSB pertaining to aircraft accident data by creating working relationships with officials from both organizations
- Ran statistical analyses on data to define a standard metric to compare the safety of different aircraft types, determined accidents per 100,000 hours of flight was most accurate and relevant; calculated this for overall safety as well as in differing environmental conditions
- Created and deployed a website in Golang to display the metrics, which were graphed using Matplotlib
- Recorded detailed methods regarding the statistical analysis to ensure reproducibility and accountability
- Created a bot to scrape the NTSB database and alert subscribed twitter users to when a final report was released
- Tech stack: Python, Jupyter Notebooks, Golang, SQL, Matplotlib

Nightingale Security *October 2017 - February 2018*
- Used the bug tracker to locate and correct various bugs in the drone control UI, such as display and logic errors
- Created last minute fixes for UX issues in time for product demonstrations to clients and venture capitalists
- Designed a feature to allow the user to create bounding boxes on images which were to be fed into a machine learning algorithm which woudl then be utilized by the drone to automatically identify risks
- Utilized agile development methods in a remote setting, including Kanban, Slack, and video conferencing
- Tech stack: JavaScript, PHP, Angular, GoogleMapsAPI, git

Topographical Data Analysis *January - April 2016*
- Reverse engineered the recording of data from the Tripod Data Systems Survey Pro in order to correct errors in manual entry of information into the total station and prevent the loss of a day's measurements
- Utilized the data from the total station to make accurate 3D images of the topography and identify systemic issues in our mapping policies including spacing and user carelessness
- Tech stack: QGIS, ROOT (CERN), Python, LaTeX

EDUCATION

Fourth Brain *2020*
Machine Learning

LambdaSchool *2017-2018*
Computer Science

Mercyhurst University
Bachelors of Science in Anthropology